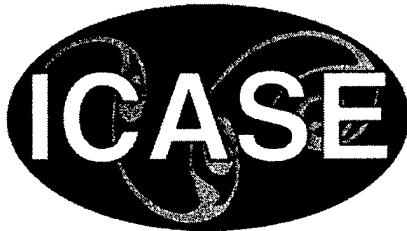NASA/CR-2000-210535
ICASE Report No. 2000-36

ICASE

On the Local Convergence of Pattern Search

Elizabeth D. Dolan
Argonne National Laboratory, Argonne, Illinois

Robert Michael Lewis
ICASE, Hampton, Virginia

Virginia Torczon
The College of William & Mary, Williamsburg, Virginia

Institute for Computer Applications in Science and Engineering
NASA Langley Research Center
Hampton, VA

Operated by Universities Space Research Association

NASA

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

September 2000

DTIC QUALITY INSPECTED 4

20001006 056

ON THE LOCAL CONVERGENCE OF PATTERN SEARCH[*]

ELIZABETH D. DOLAN[†], ROBERT MICHAEL LEWIS[‡], AND VIRGINIA TORCZON[§]

Abstract. We examine the local convergence properties of pattern search methods, complementing the previously established global convergence properties for this class of algorithms. We show that the step-length control parameter which appears in the definition of pattern search algorithms provides a reliable asymptotic measure of first-order stationarity. This gives an analytical justification for a traditional stopping criterion for pattern search methods. Using this measure of first-order stationarity, we analyze the behavior of pattern search in the neighborhood of an isolated local minimizer. We show that a recognizable subsequence converges r-linearly to the minimizer.

Key words. pattern search, local convergence analysis, global convergence analysis, desultory rate of convergence

Subject classification. Applied and numerical mathematics

1. Introduction. Pattern search methods are a class of direct search methods for solving nonlinear optimization problems. The global convergence properties of pattern search for both constrained and unconstrained problems have been established in a series of papers [7–11]. In this paper, we consider the local convergence properties of pattern search and revisit the global convergence properties in light of these new results.

For simplicity, our discussion will focus on the case of unconstrained minimization:

$$\min_{x\in\mathbf{R}^n}\ f(x).$$

Results similar to those we present here can also be derived for the general case of bound and linear constraints [9, 10]. However, the underlying ideas are simpler to explain for the unconstrained case.

We first show how the pattern size parameter, which plays a central role in the definition of pattern search methods and tacitly serves as a step-length control mechanism, also provides a reliable asymptotic measure of first-order stationarity. This gives an analytical justification for the traditional use of the pattern size parameter as a stopping criterion. We also establish a local convergence result concerning the behavior of the sequence of iterates produced by a pattern search algorithm in the neighborhood of an isolated local minimizer x_*. We show that under reasonable hypotheses the sequence of iterates converges to x_* and,

moreover, an identifiable subsequence of the iterates converges r-linearly to x_*. These analytical results are illustrated with some simple numerical experiments on quadratic objectives.

What is interesting about the analysis presented here is that we can establish local convergence properties despite the fact that direct search methods do not employ an explicit representation of the gradient of the objective and, as a consequence, cannot enforce a notion of sufficient decrease. We proved global convergence results for pattern search by showing that all iterates lie on a rational lattice. It is this restriction on the form of the steps that allows us to relax the notion of sufficient decrease and yet still prove global convergence. Pattern search may accept any point on the current lattice so long as it produces simple decrease on the value of the function at the current iterate. However, key to the global analysis is the notion of having searched in a sufficient number of directions from the current iterate to guarantee that we have not overlooked a potential direction of descent. It is only after searching over a sufficient set of directions that we are allowed to reduce the current step-length control parameter—which has the effect of refining the lattice over which we are searching.

This notion of sufficient local information at iterations at which we reduce the pattern size allows us to show that the pattern size, as measured by the step-length control parameter, provides a reliable asymptotic measure of first-order stationarity. This analytical result is gratifying since it vindicates the long-standing use of the step-length control parameter as a stopping criterion for direct search methods (see, for instance, Section 4 of [6]). The result on the correlation of the pattern size parameter and stationarity then enables us to study the local convergence properties.

Notation. We use $L(x_0)$ to denote the set $\{x \mid f(x) \leq f(x_0)\}$. It is assumed, unless otherwise noted, that all norms are Euclidean vector norms or the associated operator norm. We will use ∂ to denote the boundary of a given set. Given x and $r > 0$, we denote by $B(x, r)$ the open ball of radius r centered at x: $B(x, r) = \{\, y \mid \|\, y - x \,\| < r \,\}$.

2. Pattern search. We first review the elements of pattern search that play a role in our local analysis. There are rigorous formal definitions of pattern search [7, 11], several features of which we will shortly recall. However, pattern search can perhaps be most quickly understood with the following simple example of a pattern search algorithm. At iteration k, we have an iterate $x_k \in \mathbf{R}^n$ and a step-length control parameter $\Delta_k > 0$. Let e_j, $j = 1, \ldots, n$ be the standard unit basis vectors. For the purposes of this example, we represent the pattern of points over which we will search as the set $D \equiv \{d_i\}_{i=1}^{2n} \equiv \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$ though, as we discuss shortly, many other choices are possible. We now have several algorithmic options open to us. We will consider the simple opportunistic strategy, which is to look successively at the points $x_+ = x_k + \Delta_k d_i$, $i \in \{1, \ldots, 2n\}$ until we either find an x_+ for which $f(x_+) < f(x_k)$ or we exhaust all $2n$ possibilities. Fig. 2.1 illustrates the pattern of points among which we search for x_+ when $n = 2$.
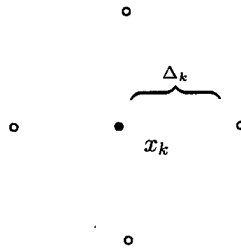

FIG. 2.1. *A simple instance of pattern search*

2

If we find no x_+ such that $f(x_+) < f(x_k)$, then we call the iteration *unsuccessful*; otherwise, we consider the iteration *successful* since we have found a new iterate that produces decrease on f at x_k. When the iteration is unsuccessful, we set $x_{k+1} = x_k$ and are required to reduce Δ_k (typically, by a half) before continuing; otherwise, for a successful iteration, we set $x_{k+1} = x_+$ and leave the step-length control parameter alone, i.e., $\Delta_{k+1} = \Delta_k$ (though the analysis also allows us to increase Δ_k if the iteration is a success). We repeat this process until some suitable stopping criterion is satisfied.

Note that overall our requirements on the outcome of the search at each iteration are light: if after searching over all the points defined by $\Delta_k d_i$, $i = 1, \ldots, 2n$ we fail to find a point $x_+ = x_k + \Delta_k d_i$ that reduces the value of f at x_k, then we must try again with a smaller value of Δ_k. Otherwise, we accept as our new iterate the first point in the pattern that produces decrease. In the latter case, we may choose to modify Δ_k. In either case, we are free to make changes to the pattern D to be used in the next iteration, though we leave the pattern unchanged in the example given above. However, in general, changes to either the step-length control parameter or the pattern are subject to certain algebraic conditions, outlined fully in [7].

A distinguishing characteristic of pattern search methods is that they sample the function over a predefined pattern of points, all of which lie on a rational lattice. By enforcing structure on the form of the points in the pattern, as well as some simple rules on both the outcome of the search and the subsequent updates, standard global convergence results can be obtained [7, 11].

There remains the question of what constitutes an acceptable pattern. A pattern must be a positive spanning set for $\mathbf{R}^n$ [4, 7]. A set of vectors $\{a_1, \ldots, a_p\}$ *positively spans* $\mathbf{R}^n$ if any vector $x \in \mathbf{R}^n$ can be written as a nonnegative linear combination of the vectors in the set; i.e.,

$$x = \alpha_1 a_1 + \cdots + \alpha_p a_p \quad \alpha_i \geq 0 \quad \forall i.$$

The set $\{a_1, \ldots, a_r\}$ is called *positively dependent* if one of the a_i's is a nonnegative combination of the others; otherwise the set is *positively independent*. A *positive basis* is a positively independent set whose positive span is $\mathbf{R}^n$.

It is straightforward to verify that the set of vectors $\{e_1, e_2, -e_1, -e_2\}$ we used to define the pattern for our simple example above is a positive spanning set, as is $\{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$ for other values of n.

2.1. Prior results. Before proceeding to our local convergence results, we recall the following proposition from [7], which we will state here without proof.

PROPOSITION 2.1. *Given any set $\{a_1, \ldots, a_r\}$ that positively spans $\mathbf{R}^n$, $a_i \neq 0$ for $i = 1, \ldots, r$, there exists $c_{2.1} > 0$ such that for all $x \in \mathbf{R}^n$, we can find an a_i for which*

$$x^T a_i \geq c_{2.1} \|x\| \, \|a_i\|.$$

Note that this is a purely geometric property of positive spanning sets.

2.2. Some formal definitions. We also need to recall some notation regarding both the pattern and the form of the search. For the details, we refer the reader to [7, 11].

We have already noted that the pattern must be a positive spanning set for $\mathbf{R}^n$. In fact, we represent the pattern using two components, a *basis matrix* and a *generating matrix*.

The basis matrix can be any nonsingular matrix $B \in \mathbf{R}^{n \times n}$.

The generating matrix is an integral matrix $C_k \in \mathbf{Z}^{n \times p_k}$, where $p_k > n + 1$. We require C_k to contain a minimum of $n + 2$ columns because the minimum number of vectors in a positive spanning set is $n + 1$ [4];

for convenience, we require a column of zeros to denote the zero step. We further partition the generating matrix so that the positive basis that guarantees that the pattern positively spans $\mathbf{R}^n$ is revealed. We call the columns associated with the positive basis the *core pattern*, which we denote as Γ_k; any remaining columns in the positive spanning set are denoted using L_k:

$$(2.1) \qquad\qquad C_k = [\quad \Gamma_k \quad L_k \quad 0 \quad].$$

We further require that $\Gamma_k \in \mathbf{\Gamma}$, where $\mathbf{\Gamma}$ comprises a finite set of integral matrices, each of which is a positive basis for $\mathbf{R}^n$.

A *pattern* P_k is then represented by the columns of the matrix $P_k = BC_k$. For convenience, we use the partition of the generating matrix C_k given in (2.1) to partition P_k as follows:

$$P_k = BC_k = [\quad B\Gamma_k \quad BL_k \quad 0 \quad].$$

To tie this notation back to the example that introduces Section 2, we note that $B = I$, $\Gamma_k = [I \ -I]$ and $L_k \equiv [0]$ (where 0 denotes the zero vector). Since the choice of Γ_k and L_k is fixed for all k, $P_k \equiv [D \ 0]$ for all k.

Now, given the step-length control parameter $\Delta_k \in \mathbf{R}$, $\Delta_k > 0$, we define a *trial step* s_k^i to be any vector of the form $s_k^i = \Delta_k B c_k^i$, where c_k^i is a column of C_k.

In Fig. 2.2 we state the general form of a pattern search method for unconstrained minimization.

Let $x_0 \in \mathbf{R}^n$ and $\Delta_0 > 0$ be given.

For $k = 0, 1, \ldots$, until convergence do:

    1. Compute $f(x_k)$.

    2. Determine a step s_k using an unconstrained exploratory moves algorithm.

    3. If $f(x_k + s_k) < f(x_k)$, then $x_{k+1} = x_k + s_k$. Otherwise $x_{k+1} = x_k$.

    4. Update C_k and Δ_k.

FIG. 2.2. *Generalized pattern search for unconstrained minimization*

We have remarkable latitude in the way in which we choose the step s_k. For the global convergence analysis to hold, we need only satisfy the hypotheses on the outcome of the unconstrained exploratory moves, given in Fig. 2.3.

---

    1. $s_k \in \Delta_k P_k$.

    2. If $\min\{f(x_k + y) \mid y \in \Delta_k B\Gamma_k\} < f(x_k)$, then $f(x_k + s_k) < f(x_k)$.

FIG. 2.3. *Hypotheses on the outcome of the unconstrained exploratory moves*

A few comments on these hypotheses are in order. The first hypothesis is straightforward: the step returned must be defined by the current pattern P_k, scaled by the current value of the step-length control parameter Δ_k. This is the condition that ensures that the steps we consider remain on the rational lattice; arbitrary steps are not allowed.

For our purposes, the second hypothesis is the more interesting. Notice that in Fig. 2.2, all that is required for a successful iteration of pattern search is that the step s_k produce simple decrease, i.e., $f(x_k + s_k) < f(x_k)$. Thus, *any* nonzero step defined by a column of $\Delta_k P_k$ that satisfies the condition $f(x_k + s_k) < f(x_k)$ may be returned by the exploratory moves since it immediately satisfies both of the hypotheses given in Fig. 2.3— even if we do not explicitly verify that $\min\{f(x_k + y) \mid y \in \Delta_k B\Gamma_k\} < f(x_k)$ is true. Thus, as we suggested

for the pattern search algorithm described in Section 2, as soon as we find a point in the pattern that satisfies this simple decrease criterion, we may terminate the iteration and declare it a success.

Recall, however, that P_k also contains a column of all zeros, so the question is: when is the zero step acceptable? The point of the second hypothesis in Fig. 2.3 is to ensure that we have sufficient information about the local behavior of f to declare an iteration unsuccessful, accept the zero step $s_k = 0$ (so that $x_{k+1} \equiv x_k$), and reduce Δ_k to continue the search with smaller steps at the next iteration. The second hypothesis implicitly decrees that we may only return the zero step when we have looked at all the steps defined by the core pattern, i.e., all steps of the form $y \in \Delta_k B\Gamma_k$. If none of the steps in the core pattern produce decrease on f at x_k, then we are free to accept the zero step. But if any step in the core pattern produces descent, the exploratory moves must return a step that produces descent—though as we have already seen, that step need not be defined by $\Delta_k B\Gamma_k$ so long as it is defined by $\Delta_k BC_k$.

For the purposes of the local convergence analysis that follows, it is the subsequence of *unsuccessful* iterates that interests us. To accept the zero step, the search must have considered all the points defined by the core pattern, which is itself a positive basis. This is where the geometric property of positive spanning sets captured in Proposition 2.1 comes in. Even though we do not have an explicit representation of $\nabla f(x_k)$ (assuming that f is differentiable), Proposition 2.1 gives us a positive lower bound, which is independent of k, on the angle between $\nabla f(x_k)$ (assuming it is nonzero) and some a_i in the positive spanning set—even though at any given iteration we do not know for which a_i this lower bound holds. However, this guaranteed lower bound, when combined with the second hypothesis in Fig. 2.3, ensures that at the end of an unsuccessful iteration, we have sufficient information about the local behavior of f at x_k. Furthermore, the quality of our local information improves as we reduce Δ_k.

Thus we have enough structure to construct local convergence results. The subsequence of unsuccessful iterates is well-defined: they are the iterations at which we can—we must—reduce Δ_k to ensure that the search can make further progress. But we reduce Δ_k only after we have sufficient local information about the behavior of f to justify this action: we have considered all the steps defined by the columns of $\Delta_k \Gamma_k$ and none of them have produced descent on f at x_k. This is the fact we now exploit.

3. Measuring first-order stationarity. The following theorem shows that the step-length control parameter Δ_k, when small enough, provides a reasonable measure of first-order stationarity when reduced after an unsuccessful iteration. For simplicity, we assume that $\nabla f(x)$ is Lipschitz continuous; however, for the reader interested in greater generality we note that a similar result can be proven under the assumption of uniform continuity.

THEOREM 3.1. *Suppose $\nabla f(x)$ is Lipschitz continuous on an open neighborhood Ω of $L(x_0)$ with Lipschitz constant C. Then there exist $\delta_{3.1} > 0$ and $c_{3.1} > 0$ for which the following holds. If x_k is an iterate at which there is an unsuccessful iteration and $\Delta_k < \delta_{3.1}$, then*

$$\| \nabla f(x_k) \| \le c_{3.1}\Delta_k.$$

Proof. Let $r = \frac{1}{2}\min\{1, \operatorname{dist}(\partial L(x_0), \partial\Omega)\}$. If $x \in L(x_0)$, then the ball $B(x, r)$ is contained in Ω. We are interested in steps of the form $s = \Delta_k Bc_k^i$, where c_k^i is a column of the core matrix Γ_k. Since $\Gamma_k \in \Gamma$ and Γ is finite, $\|s\| \le \Delta_k \|B\|\Gamma^*$, where Γ^* is the maximum norm of any column of the matrices in the set Γ. Set $\delta_{3.1} = r/(\|B\|\Gamma^*)$.

By the definition of pattern search, for any $\Gamma_k \in \Gamma$, the set $\{s \mid s \in \Delta_k B\Gamma_k\}$ forms a positive basis for

$\mathbf{R}^n$. Thus Proposition 2.1 assures us of the existence of a step s for which

$$(3.1) \qquad\qquad -\nabla f(x_k)^T s \geq c_{2.1} \|\nabla f(x_k)\| \, \|s\|.$$

Since iteration k is unsuccessful, it follows that

$$f(x_k + s) - f(x_k) \;\; \geq \;\; 0 \quad \forall \, s \in \Delta_k B\Gamma_k.$$

Since $\Delta_k < \delta_{3.1}$, $(x_k + s) \in B(x_k, r) \subset \Omega$ and we can apply the mean value theorem. In addition, using (3.1) and the Cauchy-Schwarz inequality, for some ξ in the line segment connecting x_k and $x_k + s$ we have

$$\begin{aligned}
0 &\leq f(x_k + s) - f(x_k) \\
&= \nabla f(x_k)^T s + \left(\nabla f(\xi) - \nabla f(x_k)\right)^T s \\
&\leq -c_{2.1} \|\nabla f(x_k)\| \, \|s\| + \|\nabla f(\xi) - \nabla f(x_k)\| \, \|s\|,
\end{aligned}$$

where s is the step for which (3.1) holds. Thus

$$c_{2.1} \|\nabla f(x_k)\| \leq \|\nabla f(\xi) - \nabla f(x_k)\|.$$

Again, since $B(x_k, r) \subset \Omega$, the Lipschitz continuity of $\nabla f(x)$ gives us

$$c_{2.1} \|\nabla f(x_k)\| \;\; \leq \;\; C \|\xi - x_k\| \;\; \leq \;\; C\|s\| \;\; \leq \;\; C\Delta_k \|B\| \Gamma^*.$$

Therefore

$$\|\nabla f(x_k)\| \leq c_{3.1} \Delta_k,$$

with $c_{3.1} = C\|B\|\Gamma^*/c_{2.1}$. $\square$

Theorem 3.1 gives a theoretical justification for a traditional stopping criterion for pattern search methods. In the long literature on direct search methods, one frequently encounters the suggestion that a direct search method be terminated when some measure of the step size first falls below a value deemed suitably small [2, 3, 6]. In the case of pattern search, Theorem 3.1 vindicates this intuition. At unsuccessful iterations, the step size in pattern search (as measured by Δ_k) provides a bound on first-order stationarity. At the same time, it is at the unsuccessful iterations that Δ_k is decreased. Thus, decrease in Δ_k provides a natural measure of progress which can reliably be used to test for convergence. We discuss further the use of Δ_k to measure progress when we present some numerical examples in Section 5.

A similar relation between Δ_k and constrained stationarity in the case of pattern search for bound constrained problems is explicitly used in the pattern search augmented Lagrangian algorithm in [8]. The result plays a critical role in allowing successive inexact minimization of an augmented Lagrangian without an explicit estimate of the gradient. A relation similar to Theorem 3.1 for linearly constrained pattern search appears in [10].

The global convergence analysis of pattern search in [11] says that if $L(x_0)$ is compact then we have $\liminf_{k \to \infty} \| \Delta_k \| = 0$ and $\liminf_{k \to \infty} \| \nabla f(x_k) \| = 0$. The former result and Theorem 3.1 allow us to sharpen the latter result. Let the set S represent a subsequence of unsuccessful iterations for which $\lim_{k \to \infty, k \in S} \Delta_k = 0$ (such a subsequence exists since $\liminf_{k \to \infty} \Delta_k = 0$). Then Theorem 3.1 says that we have $\lim_{k \to \infty, k \in S} \| \nabla f(x_k) \| = 0$.

The general result $\liminf_{k \to \infty} \| \nabla f(x_k) \| = 0$ for pattern search leaves open the possibility that $\| \nabla f(x_k) \|$ does not converge. In [1], Audet shows that this can actually occur by constructing a pattern search algorithm and objective for which $\{x_k\}$ has infinitely many limit points, one of which is not a

stationary point of the objective. However, Theorem 3.1 reassures us that in practice we need not worry about convergence to non-stationary points. If we stop the algorithm at the first unsuccessful iterate for which $\Delta_k < \Delta_*$ for some suitably small stopping tolerance Δ_*, then Theorem 3.1 says that $\| \nabla f(x_k) \|$ will also be small.

4. Local convergence. We next apply Theorem 3.1 to prove results on the local convergence of pattern search methods. We place the following mild hypothesis on the patterns P_k in order to bound the size of the steps s_k.

HYPOTHESIS 0. *The columns of the pattern matrix $P_k = [c_k^1 \cdots c_k^{p_k}]$ remain bounded in norm, i.e., there exists $C > 0$ such that for all k, $C > \| c_k^i \|$, for all $i = 1, \cdots, p_k$. Thus, there exists a constant $c_0 > 0$ such that any step s_k satisfies*

$$\| s_k \| \leq c_0 \Delta_k.$$

We also impose the following condition on the pattern size parameter Δ_k.

HYPOTHESIS 1. *There exists N for which Δ_k is monotonically nonincreasing for all $k \geq N$.*

Note that this is a condition we can explicitly enforce by not allowing increases in Δ_k after some iteration N; Δ_k can stay the same or decrease.

Our analysis is concerned with how pattern search behaves in the neighborhood of an isolated local minimizer x_*. We make the following assumptions about the behavior of f in the neighborhood of x_*.

HYPOTHESIS 2. *We assume that f is twice continuously differentiable on an open ball $B(x_*, \eta)$ of x_*, $\nabla f(x_*) = 0$, and that the second order sufficiency condition $\nabla^2 f(x_*) > 0$ holds at x_*.*

We further assume that $\nabla^2 f(x)$ is positive definite for all $x \in B(x_*, \eta)$. Let $\sigma_{\min}$ and $\sigma_{\max}$ be lower and upper bounds, respectively, on the singular values of $\nabla^2 f(x)$ on $B(x_*, \eta)$, and define $\kappa = \sigma_{\max}/\sigma_{\min}$.

For convenience, we will assume that $\eta < \delta_{3.1}$, where $\delta_{3.1}$ is as in Theorem 3.1. Clearly we may do this without any loss of generality. This assumption ensures that $B(x_*, \eta) \subset \Omega$ and that we may apply Theorem 3.1.

Our first result relates Δ_k to $\| x_k - x_* \|$ at unsuccessful iterates.

PROPOSITION 4.1. *Under Hypotheses 0–2, there exist $\eta > 0$, and $c_{4.1} > 0$ for which the following holds. If x_k is an iterate at which there is an unsuccessful iteration, $\Delta_k < \eta$, and $\| x_k - x_* \| < \eta$, then*

$$\| x_k - x_* \| \leq c_{4.1} \Delta_k.$$

Proof. By the mean value theorem,

$$\nabla f(x_k) - \nabla f(x_*) = \nabla^2 f(\xi)(x_k - x_*)$$

for some ξ on the line segment connecting x_k and x_*. Since $\nabla f(x_*) = 0$, we have

$$\|\nabla f(x_k)\| = \|\nabla^2 f(\xi)(x_k - x_*)\| \geq \sigma_{\min}\|x_k - x_*\|.$$

Furthermore, Theorem 3.1 holds since $\Delta_k < \eta < \delta_{3.1}$, whence

$$\sigma_{\min}\| x_k - x_* \| \leq \| \nabla f(x_k) \| \leq c_{3.1}\Delta_k.$$

Setting $c_{4.1} = c_{3.1}/\sigma_{\min}$ completes the proof. $\square$

We wish to prove that the entire sequence of iterates converges to x_*. To this end, we begin with an elementary result concerning the level sets of f near x_*.

PROPOSITION 4.2. *Under Hypothesis 2, if* $x, y \in B(x_*, \eta)$ *and* $f(x) \le f(y)$, *then*

$$(4.1) \qquad \| x - x_* \| \le \kappa^{\frac{1}{2}} \| y - x_* \|,$$

where κ *is as defined in Hypothesis 2.*

Proof. Suppose $x, y \in B(x_*, \eta)$ and $f(x) \le f(y)$. From Taylor's theorem with remainder and the fact that $\nabla f(x_*) = 0$, we have

$$f(y) = \frac{1}{2}(y - x_*)^T \nabla^2 f(\xi)(y - x_*)$$

$$f(x) = \frac{1}{2}(x - x_*)^T \nabla^2 f(\omega)(x - x_*)$$

for ξ and ω on the line segments connecting x_* with y and x, respectively. Since $f(x) \le f(y)$, we obtain

$$0 \le f(y) - f(x) = \frac{1}{2}(y - x_*)^T \nabla^2 f(\xi)(y - x_*) - \frac{1}{2}(x - x_*)^T \nabla^2 f(\omega)(x - x_*),$$

whence

$$0 \le \sigma_{\max} \| y - x_* \|^2 - \sigma_{\min} \| x - x_* \|^2,$$

and thus (4.1). $\square$

We use the previous proposition to show that if we start sufficiently close to x_* with a sufficiently small pattern size parameter Δ_k, and we stop allowing increases in Δ_k after some point (Hypothesis 1), then pattern search will not move away from a neighborhood of x_*.

PROPOSITION 4.3. *Under Hypotheses 0–2, there exist* $\delta_{4.3} > 0$, $\varepsilon_{4.3} > 0$, *and* $c_{4.3} > 0$ *for which the following holds. For* $k \ge N$, *where* N *is as defined in Hypothesis 1, if* x_k *is an iterate for which* $\Delta_k < \delta_{4.3}$ *and* $\| x_k - x_* \| < \varepsilon_{4.3}$, *then for all* $\ell \ge k$,

$$\| x_\ell - x_* \| \le \eta$$

where η *is as defined in Hypothesis 2.*

Proof. Choose $\delta_{4.3}$ and $\varepsilon_{4.3}$ to satisfy

$$\delta_{4.3} < \frac{\eta}{2c_0}$$

$$\varepsilon_{4.3} < \min \left\{ \frac{\eta}{2}, \frac{1}{2} \kappa^{-\frac{1}{2}} \eta \right\}.$$

The proof is by induction. First consider $x_{k+1} = x_k + s_k$. By Hypothesis 0, there is a constant c_0 such that $\| x_{k+1} - x_k \| = \| s_k \| < c_0 \Delta_k$. We have, *a priori*,

$$\| x_{k+1} - x_* \| \le \| x_{k+1} - x_k \| + \| x_k - x_* \| < c_0 \Delta_k + \varepsilon_{4.3} < \eta,$$

so $x_{k+1} \in B(x_*, \eta)$. Since $f(x_{k+1}) \le f(x_k)$, and $x_{k+1} \in B(x_*, \eta)$, from Proposition 4.2 we obtain

$$\| x_{k+1} - x_* \| \quad \le \quad \kappa^{\frac{1}{2}} \| x_k - x_* \| \quad \le \quad \kappa^{\frac{1}{2}} \varepsilon_{4.3} \quad < \quad \eta.$$

Now consider any $\ell \ge k + 1$, and suppose

$$\| x_\ell - x_* \| \le \eta.$$

8

Then

(4.2)
$$\| x_{\ell+1} - x_* \| \le \| x_{\ell+1} - x_\ell \| + \| x_\ell - x_* \|.$$

Hypothesis 1, assures us that $\Delta_\ell \le \Delta_k$ for $\ell \ge k$, so

$$\| x_{\ell+1} - x_\ell \| \le c_0 \Delta_\ell \le c_0 \Delta_k.$$

Meanwhile, by the induction hypothesis, $x_\ell \in B(x_*, \eta)$. Since $f(x_\ell) \le f(x_k)$, as well, Proposition 4.2 and the assumption $\| x_k - x_* \| < \varepsilon_{4.3}$ say that

$$\| x_\ell - x_* \| \le \kappa^{\frac{1}{2}} \| x_k - x_* \| \le \kappa^{\frac{1}{2}} \varepsilon_{4.3}.$$

Thus (4.2) yields

$$\| x_{\ell+1} - x_* \| < c_0 \Delta_k + \kappa^{\frac{1}{2}} \varepsilon_{4.3} < \eta.$$

□

For the final local convergence result, we need the following additional hypothesis, which underlies the global convergence analysis of pattern search [11].

HYPOTHESIS 3. *The set $L(x_0)$ is compact.*

From [11] we know that if $L(x_0)$ is compact, $\liminf_{k \to \infty} \Delta_k = 0$. Thus, Hypothesis 1 and Hypothesis 3 together mean that $\lim_{k \to \infty} \Delta_k = 0$.

Putting the pieces together, we obtain the following local convergence result for pattern search.

THEOREM 4.4. *Suppose Hypothesis 3 holds. Given a pattern search algorithm satisfying Hypotheses 0–1, suppose there exists a limit point x_* of the sequence of iterates $\{x_k\}$ produced by the algorithm that is a local minimizer satisfying Hypothesis 2.*

Then there exist $c_{4.4} > 0$ and K such that for all $k \ge K$,

(4.3)
$$\| x_k - x_* \| \le c_{4.4} \Delta_{m(k)},$$

where $m(k)$ is the last unsuccessful iterate preceding k. As a consequence, we have $\lim_{k \to \infty} x_k = x_*$.

Proof. Recall N from Hypothesis 1, the iteration after which we no longer allow Δ_k to increase. Since x_* is a limit point of $\{x_k\}$ and $\lim_{k \to \infty} \Delta_k = 0$ (by Hypotheses 1 and 3), there exists an iterate x_{k_1}, $k_1 \ge N$ such that

$$\Delta_{k_1} < \min\{\eta, \delta_{4.3}\}$$
$$\| x_{k_1} - x_* \| < \min\{\eta, \varepsilon_{4.3}\},$$

where η is as in Hypothesis 2 and $\delta_{4.3}$, $\varepsilon_{4.3}$ are as in Proposition 4.3.

By Proposition 4.3, we have

$$\| x_k - x_* \| \le \eta$$

for all $k \ge k_1$. Then, by Proposition 4.1, we have

$$\| x_m - x_* \| \le c_{4.1} \Delta_m$$

for all unsuccessful iterates x_m with $m \ge k_1$.

9

Now let K be the first iteration after k_1 at which we have an unsuccessful iteration. For all unsuccessful iterations $k \geq K$, Proposition 4.1 gives us

$$(4.4) \qquad \| x_k - x_* \| \quad \leq \quad c_{4.1}\Delta_k \quad = \quad c_{4.1}\Delta_{m(k)},$$

since $k = m(k)$ in this case. Meanwhile, for the successful iterations $k \geq K$, we have $f(x_k) < f(x_{m(k)})$, and Proposition 4.3 assures us that $\| x_k - x_* \| \leq \eta$. It then follows from Proposition 4.2 that

$$(4.5) \qquad \| x_k - x_* \| \quad \leq \quad \kappa^{\frac{1}{2}}\| x_{m(k)} - x_* \| \quad \leq \quad \kappa^{\frac{1}{2}}c_{4.1}\Delta_{m(k)}.$$

Then (4.3) follows from (4.4) and (4.5), with $c_{4.4} = \kappa^{\frac{1}{2}}c_{4.1}$.

Since Hypotheses 1 and 3 imply $\Delta_k \to 0$, it follows that $\lim_{k\to\infty} x_k = x_*$. $\square$

Theorem 4.4 complements Theorem 3.7 of [11], where it is shown, under different hypotheses and a more stringent criterion for accepting a step, that $\| \nabla f(x_k) \| \to 0$.

Theorem 4.4 says that for the *subsequence* of unsuccessful iterates, the rate of convergence is r-linear. Theorem 4.4 says nothing about what may happen at the successful iterations, nor how many such iterations there may be between unsuccessful iterations. What we have is a sort of multi-step r-linear rate of convergence, but one for which we do not know, and, as our numerical tests indicate, cannot predict, what the number of intervening steps may be. For want of an existing term for this notion of convergence, we call it *desultory r-linear convergence*.

The obstruction to sharpening the rate of convergence result is that we do not know *a priori* how much improvement we obtain in $f(x)$ at the successful iterations. At unsuccessful iterates, we reach the decision that we can only make progress by reducing Δ_k only after we evaluate $f(x)$ along a set of directions that necessarily includes a descent direction (see Fig. 2.3 and Proposition 2.1). On the other hand, successful iterates are less informative. We have no *a priori* idea of how much improvement we obtain in a successful iteration relative to the minimization of the quadratic model of $f(x)$ along the pattern search directions. Given the paucity of information generally available in pattern search one must not expect too much.

More positively, Theorem 4.4 suggests how one can "accelerate" the local convergence of pattern search algorithms. One need only rename the formerly unsuccessful iterates successful iterates and drop the formerly successful iterates from discussion. Then, *mirabile dictu*, this simple modification makes the successful iterations an r-linearly convergent sequence.

5. Numerical results. We now present some numerical experiments that illustrate the practical implications of the local convergence results. The first round of testing, reported in Section 5.2, supports the analysis; the second round, reported in Section 5.2, shows its limitations.

The numerical results we report are in no way exhaustive, and simply serve to illustrate how the local convergence results are manifest in practice. The results we report here on the effectiveness of Δ_k as a measure of stationarity are, in part, a summary of some of the results reported in [5]. The second round of testing regarding the local rate of convergence was based on the implementation of pattern search developed for, and reported in, [5].

5.1. The testing environment. Full details of the numerical experiments can be found in [5]. The tests we report here were done with randomly generated positive definite quadratic functions. This is a reasonable choice, since we are interested in the local convergence behavior of pattern search, and any C^2 function looks like a convex quadratic in the neighborhood of an isolated local minimizer. The quadratics tested were of the form $f(x) = x^T A x + c$, where $A = H^T H$ and $H \in \mathbf{R}^{(n+2)\times n}$ is a matrix with entries

that are normal random variates with means of zero and standard deviations of one. The lack of a linear term causes the solution to lie at the origin. The constant term c is not interesting for the purposes of the optimization, but provides a useful tag for identifying individual functions. For the testing in [5], $2 \leq n \leq 5$; we show two results for $n = 5$.

In addition to randomly generating the entries of the matrix H, we also randomly generated Δ_0 and the entries of the vector x_0. The entries for the starting point x_0 were also normal random variates with means of zero and standard deviations of one. The choice for Δ_0 was an exponential variate with a mean of one.

The software described in [5] was written in C++ to make use of C++ classes, a convenient way to establish the key features of pattern search and then easily derive specific variants. Several of these variants were implemented and tested, as described in [5]. We show results using HJSearch, an implementation of the classical pattern search algorithm of Hooke and Jeeves [6]; CompassSearch, the pattern search algorithm described in Section 2; and NLessSearch, a pattern search algorithm that takes advantage of the fact that a minimal positive basis requires only $n + 1$ vectors [7], as opposed to the $2n$ coordinate vectors used in most traditional pattern search methods, including compass search and Hooke and Jeeves.

5.2. Measuring stationarity. The first question we ask is how effective is Δ_k as a measure of stationarity? Not too surprisingly, the results of our tests showed that Δ_k is a reliable measure of progress toward a solution. Furthermore, our numbers make quite clear the r-linear behavior of the sequence of unsuccessful iterates.

After any unsuccessful iteration, a pattern search method is required to reduce Δ_k. We used the standard reduction factor of $\frac{1}{2}$ so that after an unsuccessful iteration, $\Delta_{k+1} = \frac{1}{2}\Delta_k$. Before proceeding to the next iteration, we recorded the value of Δ_k, $\|\nabla f(x_k)\|$, $|f(x_k) - f(x_*)|$, and $\|x_k - x_*\|$ (though since we knew $x_* \equiv 0$, we simply had to compute $\|x_k\|$). Representative results from two particular tests are given in Tables 5.1 and 5.2.

TABLE 5.1
NLessSearch *in five variables*

Δ_k	$\|\nabla f(x_k)\|$	$|f(x_k) - f(x_*)|$	$\|x_k - x_*\|$
0.696226813823902	3.868198250940641	0.97865190917639	0.277727609425140
0.348113406911951	1.774568869170108	0.29434050891286	0.217654912583381
0.174056703455976	1.453229516907024	0.14479398150378	0.108469794156683
0.087028351727988	0.281730564593278	0.02631186626609	0.115903974197448
0.043514175863994	0.294368401951189	0.01457817127405	0.070417918465458
0.021757087931997	0.139178781830330	0.00462725481091	0.040571903721886
0.010878543965999	0.029890455941604	0.00073483024738	0.034760919692137
0.005439271982999	0.029890455941604	0.00073483024738	0.034760919692137
0.002719635991500	0.026442509744175	0.00033346616931	0.024811502814971
0.001359817995750	0.012307380822946	0.00022911316873	0.021849036278915
0.000679908997875	0.009511207300009	0.00014689251267	0.017596624227294
0.000339954498938	0.005426023491341	0.00004796415094	0.010013823809977
0.000169977249469	0.002065061853736	0.00000720339894	0.003890485486089
0.000084988624734	0.001038174548727	0.00000179297966	0.001938237809581
0.000042494312367	0.000435621809509	0.00000034070778	0.000849356148369
0.000021247156184	0.000231605471572	0.00000008825774	0.000432171965702
0.000010623578092	0.000113220891634	0.00000001411259	0.000168890905210
0.000005311789046	0.000067613953149	0.00000000577676	0.000108943399928

TABLE 5.2

HJSearch *in five variables*

Δ_k	$\|\nabla f(x_k)\|$	$|f(x_k) - f(x_*)|$	$\|x_k - x_*\|$
0.696226813823902	3.718628968450993	3.96639084353257	2.396301558944381
0.348113406911951	1.370661155865317	0.44618879006458	0.698389592313846
0.174056703455976	0.993386770046628	0.19091214014793	0.450386903073632
0.087028351727988	0.236893510661273	0.01477525409286	0.153943970082610
0.043514175863994	0.314026005456998	0.01421309666224	0.119315177505950
0.021757087931997	0.131650296045321	0.00223337373949	0.034002609804365
0.010878543965999	0.042526372212693	0.00028996796577	0.015791910616849
0.005439271982999	0.032921235371376	0.00018078437086	0.014678346820778
0.002719635991500	0.012854930063180	0.00014567060113	0.016582990396810
0.001359817995750	0.005667414556147	0.00001023084696	0.003757596625046
0.000679908997875	0.004101406209192	0.00000429756349	0.002612852810391
0.000339954498938	0.001396318029208	0.00000050775161	0.000854609846084
0.000169977249469	0.000833651146770	0.00000049903818	0.000985750547712
0.000084988624734	0.000563050121378	0.00000002356890	0.000074244150774
0.000042494312367	0.000112117511534	0.00000000325088	0.000043510325021
0.000021247156184	0.000097664692564	0.00000000266601	0.000032689236837
0.000010623578092	0.000035578092711	0.00000000026108	0.000013878637584
0.000005311789046	0.000010624256256	0.00000000015362	0.000017458315183

The purpose of the results we report in Tables 5.1 and 5.2 is not to demand close scrutiny of each entry, but rather to demonstrate the trends in each of the four quantities measured. We clearly see the r-linear behavior the analysis tells us to expect: by the time we halve Δ_k, we have roughly halved the error in the solution. (Note that in Table 5.1 we have two consecutive unsuccessful iterations.)

We report here the results from only two experiments, but these are representative of results from ten thousand runs over multiple quadratics, in multiple dimensions, from multiple starting points, with multiple choices of Δ_0, using four different pattern search methods. We found that across all these tests Δ_k gave us a consistent measure of the accuracy of the solution.

One beauty of using Δ_k as a measure of stationarity is that it is perfectly natural in the context of pattern search; no additional computation is required. Another strong reason for using Δ_k as a measure of stationarity is that it is remarkably insusceptible to error. Once an initial choice of Δ_k has been made, the only possible numerical error occurs when converting the decimal representation to binary. Since most implementations use the value of $\frac{1}{2}$ to reduce Δ_k, which requires only a simple binary shift, there are no additional numerical issues to cloud the computation of this measure. We close with one final observation to be made about the practical utility of Δ_k as a measure of stationarity. A useful feature of pattern search is that one only requires ranking, or order, information to drive the search—no numeric values for the function are necessary [7]. In such a setting, Δ_k is a feasible measure of progress whereas measures based on the numeric values of the objective function are not.

6. How many successful iterates? Theorem 4.4 says that the subsequence of unsuccessful iterates converges r-linearly once we are in the neighborhood of a solution. A natural question to then ask is: how many iterations occur in practice between each iterate included in this subsequence? This question is more subtle, and our results are less conclusive. Again, we illustrate our findings with only a few specific examples in Figs. 6.2–6.5.

In all instances, we give a final stopping tolerance of 2×10^{-8}; i.e., we terminate the search when $\Delta_k < 2 \times 10^{-8}$. Along the horizontal axis, we list the number of unsuccessful iterations; i.e., the number of times we halve Δ_k before it is less than the stopping tolerance. Each bar then represents the number of successful iterations that preceded an unsuccessful iteration plus the (single) unsuccessful iteration so that summing all the entries gives us the total number of iterations for the search.

Notice that for the three algorithms we tested, the scale on the vertical axes varies considerably. For NLessSearch, the number of successful iterations preceding an unsuccessful iteration can be considerably higher than, say, for HJSearch, but overall, the results are inconclusive. We cannot predict how many successful iterations may proceed an unsuccessful iteration, nor does there seem to be any particular trend.

The only substantive observation we can extract from our tests is illustrated by the example shown in Fig. 6.1.
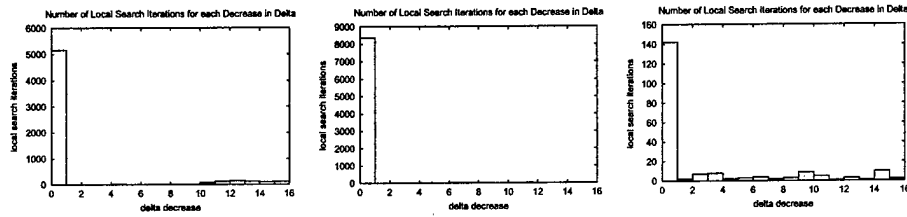

FIG. 6.1. *NLess, Compass, & HJ in 4 variables*

Here the explanation for the huge number of successful iterations before Δ_k is ever reduced seems to be straightforward. The initial choice of Δ_0, which is drawn randomly, is so small (0.001128116614106) that initially there is a long sequence of successful iterations, but progress is remarkably slow because we start with such a small choice of Δ_0 that all the trial steps are quite short. The obvious conjecture to make is that it is better to start with a reasonably large choice of Δ_0—a practical issue that is the subject of another study in preparation.
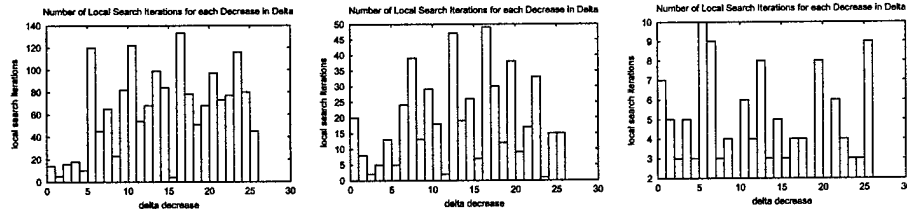

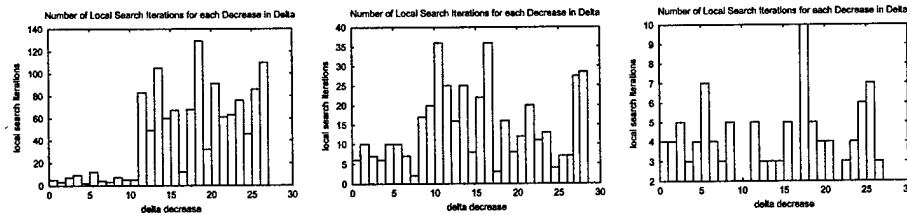FIG. 6.2. *NLess, Compass, & HJ in 8 variables*
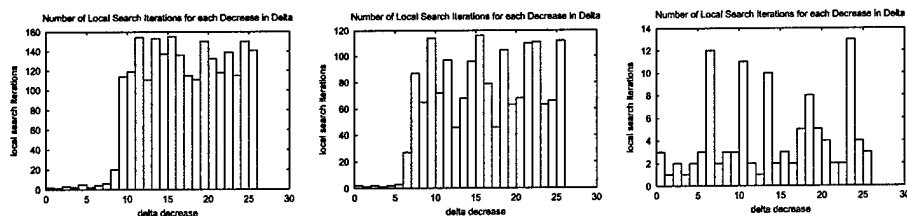

FIG. 6.3. *NLess, Compass, & HJ in 8 variables*
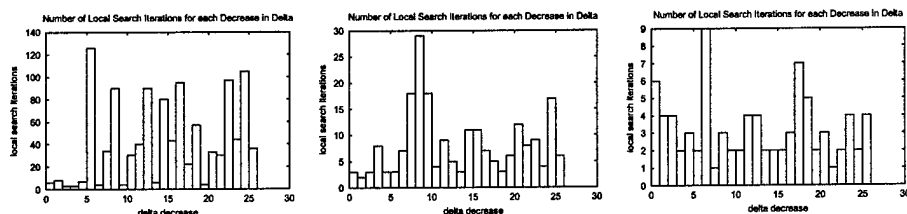
FIG. 6.4. *NLess, Compass, & HJ in 4 variables*

FIG. 6.5. *NLess, Compass, & HJ in 4 variables*

7. Conclusion. The results given here round out the convergence analysis of pattern search. The analysis and numerical experiments reported here show that Δ_k can be used as a reliable stopping criterion. Moreover, these tests show that the correlations predicted by Theorems 3.1 and 4.4 between Δ_k, $\| \nabla f(x_k) \|$, and $\| x_k - x_* \|$ are manifest in practice. These results vindicate the intuition of the early developers of direct search methods.

REFERENCES

[1] C. AUDET, *Convergence results for pattern search algorithms are tight*, Tech. Rep. 98–24, Department of Computational and Applied Mathematics, Rice University, 6100 Main Street, MS 134, Houston, Texas 77005–1892, 1998.

[2] M. AVRIEL, *Nonlinear Programming: Analysis and Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1976.

[3] M. J. BOX, D. DAVIES, AND W. H. SWANN, *Non-Linear Optimization Techniques*, ICI Monograph No. 5, Oliver & Boyd, Edinburgh, 1969.

[4] C. DAVIS, *Theory of positive linear dependence*, American Journal of Mathematics, 76 (1954), pp. 733–746.

[5] E. D. DOLAN, *Pattern search behavior in nonlinear optimization.* Honors Thesis, Department of Computer Science, College of William & Mary, Williamsburg, Virginia 23187–8795, May 1999. Accepted with Highest Honors.

[6] R. HOOKE AND T. A. JEEVES, *Direct search solution of numerical and statistical problems*, Journal of the Association for Computing Machinery, 8 (1961), pp. 212–229.

[7] R. M. LEWIS AND V. J. TORCZON, *Rank ordering and positive bases in pattern search algorithms*, Tech. Rep. 96–71, Institute for Computer Applications in Science and Engineering, Mail Stop 132C,

NASA Langley Research Center, Hampton, Virginia 23681–2199, 1996. In revision for Mathematical Programming.

[8] ——, *A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds*, Tech. Rep. 98–31, Institute for Computer Applications in Science and Engineering, Mail Stop 132C, NASA Langley Research Center, Hampton, VA 23681–2199, July 1998. In review for SIAM Journal on Optimization.

[9] ——, *Pattern search algorithms for bound constrained minimization*, SIAM Journal on Optimization, 9 (1999), pp. 1082–1099.

[10] ——, *Pattern search methods for linearly constrained minimization*, SIAM Journal on Optimization, 10 (2000), pp. 917–941.

[11] V. TORCZON, *On the convergence of pattern search algorithms*, SIAM Journal on Optimization, 7 (1997), pp. 1–25.